Engineering Educators Bringing the World Together
2025 ASEE Annual Conference & Exposition
Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025
Paper ID #46936

BOARD # 373: HDR DSC: Interactive data science education for civil engineers

Prof. David Lattanzi, George Mason University

Dr. David Lattanzi is an associate professor and the John Toups Faculty Fellow in the Sid and Reva Dewberry Department of Civil, Environmental, and Infrastructure Engineering at George Mason University. He studies how robotics and artificial intelligence technologies can be developed to improve engineering design and evaluation processes. Dr. Lattanzi is a licensed professional engineer with experience in both the design and inspection of complex bridge structures.

HDR DSC: Interactive data science education for civil engineers

Introduction

Applications of artificial intelligence and data science for civil engineering are growing rapidly. In particular, the need for enhanced community resilience in the face of climate change and related natural disasters has become central to the practice of civil engineering, and the rapid data analysis and decision making afforded by data science methods is essential to this task. The related professional focus on sustainability also means that there is increased pressure on engineers to optimize complex system designs through data-driven approaches. In tandem, data analytics programming interfaces and software applications are now accessible to those without advanced programming and data science skills. While these new technologies expand the pool of data science users, they also create new risks related to a lack of understanding of the fundamentals of these systems. A direct analogue for the current dilemma can be found in the early implementations of finite element analysis (FEA). Misuse and misunderstanding of FEA methods led to engineering failures across a variety of domains. For example, misunderstanding of FEA modeling concepts likely played a central role in the collapse of the Hartford Civic Center and the sinking of the Sleipner A offshore platform [1], [2].

Data science education for civil engineering students poses several unique challenges. The first is that modern civil engineering education heavily focuses on physics-driven analysis and modeling methods, using coursework in fundamental concepts to provide the foundation for more advanced course topics. For example, a structural engineering student will receive coursework in statics, mechanics of materials, and structural analysis before taking a course in finite elements. A computer science student taking a course in machine learning would see a similar set of prerequisites, but it is impractical to force these requirements on civil engineering students that still need training in conventional topics. Additional pressures come from the fact that traditional engineering skills are the basis for most licensure requirements and are the skills most sought after by future employers looking for entry-level employees.

Additionally, there are significant pedagogical challenges. Civil engineering students interested in data science often take the related coursework as electives in other departments. While this reduces the course load burden for faculty, it creates a situation where students are learning the fundamental data science concepts without contextualization. The educational literature has demonstrated the negative consequences of decontextualization with respect to content retention and student enthusiasm [3]. However, it is non-trivial for civil engineering programs to offer integrated data science courses, as it can be challenging to find a viable instructor with both the necessary training and ability to present content in a civil engineering context.

Focus of this work

The focus of this work is to develop an approach to data science education for civil engineers that addresses curricular restraints, the need for contextualization, and the need for specialized instructor training. The proposed solution is to integrate data science directly into existing course offerings through self-contained interactive modules. Rather than reorganize curriculum as was done for FEA, data science topics are taught alongside fundamental engineering topics in a way that strengthens understanding of both.

Presented here is one component of a larger initiative devoted to this effort, in the context of a widely offered undergraduate course, Mechanics of Materials. Key data science concepts were first identified. Then, an evaluation of the existing curriculum identified opportunities to teach each concept within the context of a core Mechanics topic. A series of self-contained interactive programming modules were then developed, with an emphasis on exploratory learning and a focus on reducing or eliminating the need for programming expertise by the student. These modules were then piloted and evaluated within the context of a semester-long Mechanics course.

Module design

Each module needed to include interactivity and exploratory thinking to the extent possible, in order to foster curiosity in the student to engage in a non-traditional topic. Each module needed to clearly introduce the relevant data science topic and directly connect it to civil engineering. Finally, each module included reflection questions for the student to encourage critical thinking about the data science topic. Two modules were designed and implemented as MATLAB [4] live notebooks that allowed the integration of instructional content and dynamic interactive programming exercises. A third module was more open-ended, as will be discussed.

Tension testing was the first opportunity identified. The tension test module is designed to extend on conventional stress-strain curve parameterization by first introducing bi-linear models (Figure 1). The concept of polynomial regression and error minimization are then introduced, and the student is encouraged to explore how polynomial model order impacts curve fitting. Neural networks are then discussed as an alternative to polynomial regression that does not require the user to assume a model order. The student is then encouraged to explore how simple network architectures fit to the data.

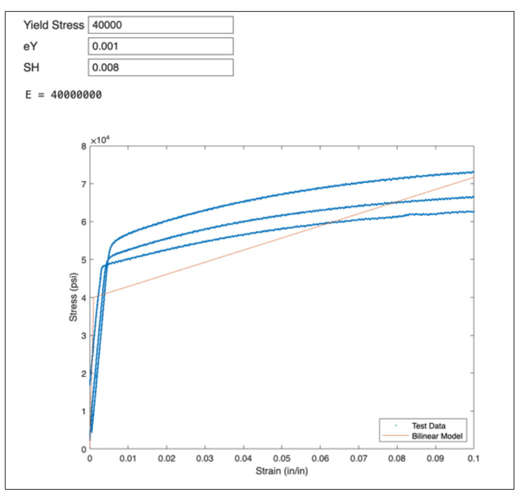
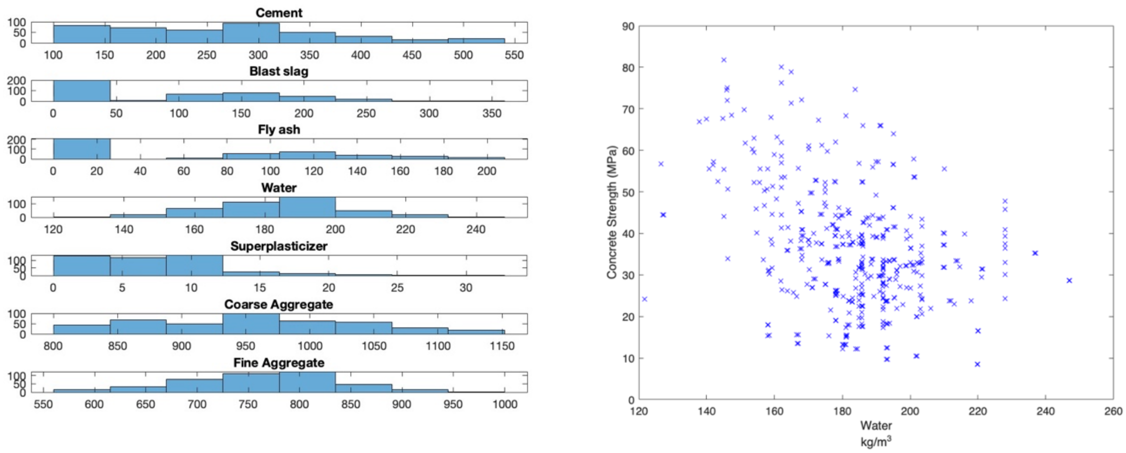


Figure 1. Interactive bilinear model fitting exercise

The behavior of civil engineering materials was the second module. For this initiative, the properties of plain cement concrete, and their correlation with compressive strength, were selected as a basis for teaching the concept of exploratory data analysis (EDA) [5]. EDA is the process of exploring a dataset to understand its characteristics and potential for data science applications. This typically involves understanding the data distributions, normality, separability, and strong correlations within the dataset. For this module, students were shown the process of EDA using the canonical "Iris Setosa" dataset, and were then encouraged to explore the concept using a dataset of concrete mix designs (Figure 2).



(A) (B)

Figure 2. Interactive exploratory data analysis on the properties of concrete, and their correlation with concrete compressive strength. A) histograms showing admixture quantities in kg/m^3 on the x-axis against their frequency of occurrence on the y-axis. The number of bins is controllable by the user to allow visualization of distribution parameters. B) User driven visualization of the correlations between mixture quantities and concrete compressive strength.

The third educational module focused on Generative AI. For this module, the students were first instructed to use a Large Language Model (LLM) chat client (e.g. Chat-GPT or Bard) to teach themselves about the fundamentals of Generative AI. They then needed to select a Mechanics course topic and ask Presentations.AI [6], a generative AI tool, to create a slide show presentation on their topic. The follow up questions instruct the student to critically assess the presentation to see what the AI model did well, what it missed, and why it produced the results that it did.

Evaluation

The learning modules were offered as extra credit, in order to create a control group of students that did not do the exercises. Of the 38 students in the class, 28 chose to do assignments and 10 students declined. The learning modules were evaluated in two ways. The students that participated in the learning modules were asked to fill out a survey at the end of the semester (Figures 3-5).

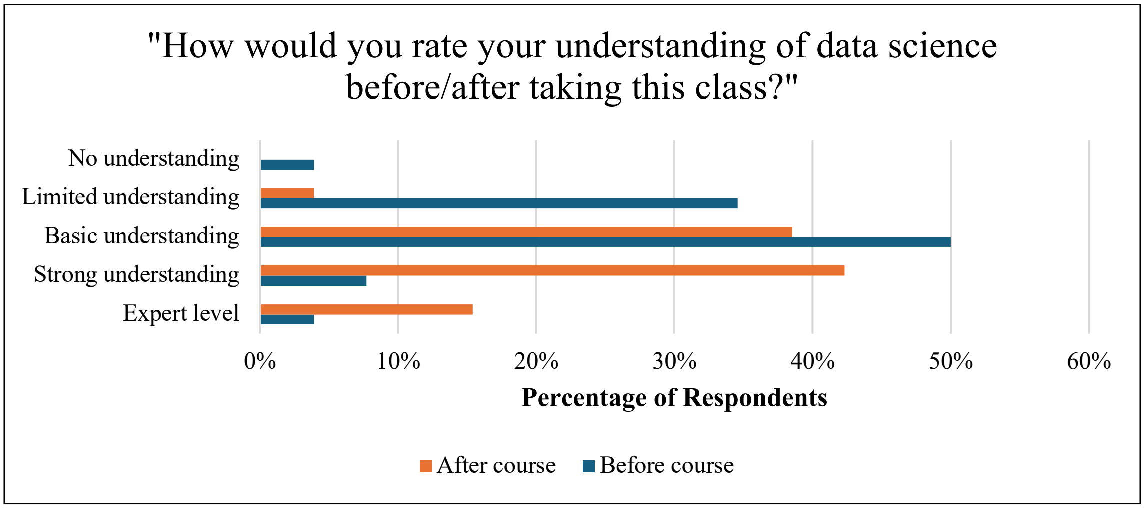


Figure 3. Student responses regarding understanding of data science before and after the course.

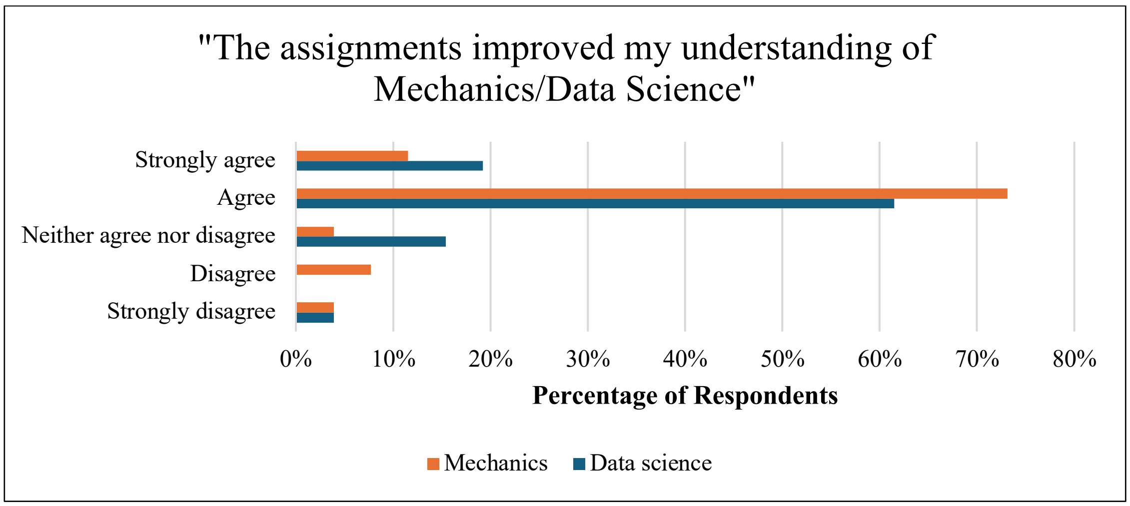


Figure 4. Student responses on the impact of the interactive modules on their understanding of both core course concepts (Mechanics) and data science.

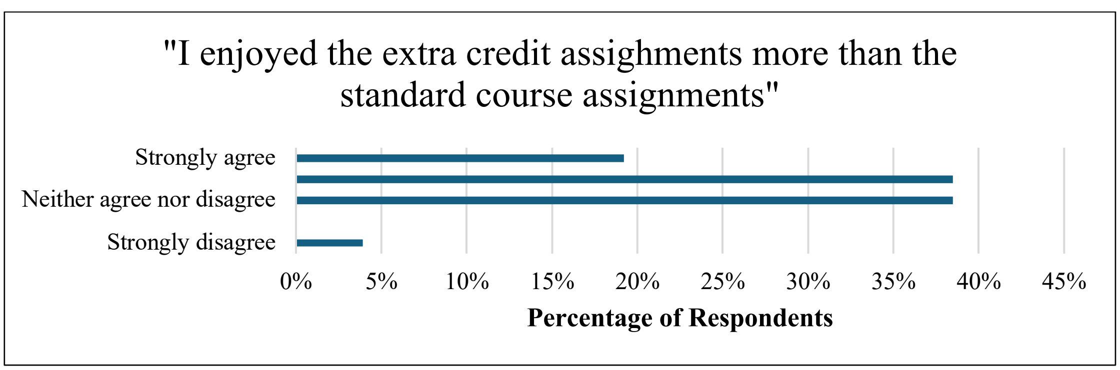


Figure 4. Student responses regarding enjoying of the supplemental exercises.

Additional evaluation involved tracking of performance indicators related to the civil engineering topics presented in the modules. These indicators included a homework assignment, a midterm exam question, and the student's final overall grade. An analysis of these indicators was inconclusive. The homework scores and midterm scores were on average higher for the control group, despite student comments suggesting that the learning modules helped their understanding of core topics. The most likely explanation for this is that students that opted out of the assignments were students with overall higher grade averages, with an average final grade of 88.9%, compared to an average of 82.4% for the students that did the assignments.

Conclusion

The pilot implementation of these course modules was clearly successful. The majority of participants enjoyed the modules, often more than the core class content, and felt that they were informative. Perhaps the most interesting survey response was that over 84% of students felt that the modules improved their understanding of fundamental civil engineering topics. While this was not reflected in the analysis of performance indicators, that is most likely an issue with the evaluation process. For the next implementation, a different approach to performance analysis is warranted.

Future work will investigate how to integrate these concepts into other relevant classes. Additionally, the modules will be refined in order to reduce the technical challenges of working with MATLAB. Eventually, these modules will be openly shared and evaluated at different institutions nationwide.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 2123237.

References

[1] R. Martin and N. J. Delatte, “Another Look at Hartford Civic Center Coliseum Collapse,” *J. Perform. Constr. Facil.*, vol. 15, no. 1, pp. 31–36, Feb. 2001, doi: 10.1061/(ASCE)0887-3828(2001)15:1(31).
[2] B. Jakobsen, “The Sleipner accident and its causes,” *Eng. Fail. Anal.*, vol. 1, no. 3, pp. 193–199, Oct. 1994, doi: 10.1016/1350-6307(94)90018-3.
[3] M. S. Kleine, K. Zacharias, and D. Ozkan, “Contextualization in engineering education: A scoping literature review,” *J. Eng. Educ.*, vol. 113, no. 4, pp. 894–918, 2024, doi: 10.1002/jee.20570.
[4] “MATLAB.” Accessed: Jan. 15, 2025. [Online]. Available: https://www.mathworks.com/products/matlab.html
[5] V. Cox, “Exploratory Data Analysis,” in *Translating Statistics to Make Decisions*, Apress, Berkeley, CA, 2017, pp. 47–74. doi: 10.1007/978-1-4842-2256-0_3.
[6] “Presentations.AI - ChatGPT for Presentations.” Accessed: Jan. 15, 2025. [Online]. Available: https://www.presentations.ai/